Phonetica
Editors: H. PILCH, Freiburg i. Br.; G. UNGEHEUER, Bonn
Publishers: S. KARGER, Basel
SEPARATUM (Printed in Switzerland)

Phonetica 34: 295–303 (1977)

Laryngeal Timing in Consonant Distinctions

ARTHUR S. ABRAMSON

University of Connecticut, Storrs, Conn., and Haskins Laboratories, New Haven, Conn.

The timing of the valvular action of the larynx may be said to be a physiological mechanism that underlies such acoustic phonetic features as the onset and offset of voice pulsing, intensity of plosive release, amount of aspiration noise, attenuation of the first formant, onset of voice-excited formant transitions, and perturbations of fundamental frequency. These features intersect in various combinations to furnish the phonetic basis of phonologically relevant voicing and aspiration[1]. They also seem to cover most instances of the vaguely defined term 'tense' or 'fortis' as applied to consonants[2].

In our early approach to these matters [LISKER and ABRAMSON, 1964, 1965; ABRAMSON and LISKER, 1965][3], LEIGH LISKER and I focused our attention on stop-consonant distinctions in word-initial position. For our cross-language investigations, this choice made sense, because the richest sets of contrasts are most often found in initial stops. We hypothesized that temporal variations in glottal settings for phonation would differentiate most homorganic consonants said to be distinguished phonologically by such features as voicing, aspiration, and tensity. Since, in those days – and to a great extent to this day – it was difficult to make extensive physiological observations of the action of the larynx, we used instrumental displays of the acoustic signal for analysis. The most convenient acoustic index to the closing of the glottis for

[1] The phonemic use of voiced aspiration is not fully handled by laryngeal timing alone; it also requires a dimension of glottal aperture.

[2] e.g., in English or Spanish.

[3] In this short review, I shall cite mainly work done in collaboration with a few of my colleagues. Certain references needed to document controversial matters and theoretical points will also be given.

phonation in initial position was the beginning of regular vertical striations corresponding in a wide-band spectrogram to the quasi-periodic voice pulses of speech. We proposed the term voice onset time (VOT) which we defined as the temporal relation between the onset of glottal pulsing and the release of the initial stop consonant. Specifically, voicing detected before the release, that is, during the stop occlusion, was called voicing lead, while voicing starting after the release was called voicing lag.

By and large, we found that VOT is indeed a very good index to laryngeal timing for the types of homorganic stop consonants in question. The measure provided rather good separation for labial, dental, alveolar, retroflex, and velar stops across a variety of languages that have two or three distinct classes at each place of articulation [LISKER and ABRAMSON, 1964, 1967][4]. Adopting the convention of assigning a timing value of zero to the moment of stop release, negative values to voicing lead, and positive values to voicing lag, we found an essentially tri-modal distribution of VOT values for eleven languages that were examined. The first mode centers at –100 msec for a range of values representing voiced unaspirated stops. The second mode centers at +10 msec and corresponds most generally to voiceless unaspirated stops. The third mode centers at +75 msec and corresponds to voiceless aspirated stops. Voicing lag, seemingly occurring for the most part with an open glottis, was regularly accompanied by turbulent excitation of the upper vocal tract (aspiration); in addition, attenuation of the first formant was often visible in the spectrogram [i.e., 'F1 cutback' [LIBERMAN *et al.*, 1958]].

It is clear then that VOT is not an acoustic continuum, although it may be viewed as an articulatory or physiological continuum. In using techniques of speech synthesis to validate our acoustic findings, we varied values of voicing lead and voicing lag with the latter including increments of cutback of the first formant and noise excitation of the upper formants. With stimuli simulating labial, apical, and dorsal CV syllables, we demonstrated the perceptual efficacy of the VOT dimension across a few languages [ABRAMSON and LISKER, 1965, 1970a, 1973; LISKER and ABRAMSON, 1970]. Since this work has stimulated many studies on the part of others, gratifyingly too numerous to list

[4] For those with a fourth laryngeal class, see footnote 1. Ejectives, not considered here, may also be said to involve laryngeal timing; however, it is the timing of the tight closing of the vocal folds relative to oral closure that is relevant.

here, it is important to stress that psychological and linguistic discussions of VOT should not give the impression that it is an acoustically simple dimension. It is radically different from many other continua in the literature in that there is an abrupt qualitative discontinuity at the point of stop release. Discussions of special mechanisms for the processing of speech, feature detectors, and other related matters must make it clear that we are dealing with profound psychoacoustic shifts[5]. Voicing lead presents the ear with a low-amplitude, low-frequency spectrum during the initial part of the stimulus. In the absence of lead, we have the sudden full unfolding of the formant pattern for the syllable. For appreciable values of voicing lag, the noise excitation of the formant pattern with its sudden shift to a train of voicing pulses has been shown by our data to be psychoacoustically easier to process.

I fear that our coining of the term voice onset time with its popular acronym VOT, handy as it was for much of our research, has led some colleagues in the field astray. A more appropriate concept is simply that of voice timing – i.e. laryngeal timing – which subsumes VOT as a special case. Some scholars finding VOT very useful for their purely perceptual speculations have perhaps found little interest in our more physiological endeavors, which, I think, put our acoustic and perceptual data into proper perspective. Transillumination of the larynx [Lisker *et al.*, 1969], fiberoptic observations [Sawashima *et al.*, 1970; Lisker *et al.*, 1970; Cooper *et al.*, 1971], and electromyographic recordings combined with fiberoptic observations [Hirose *et al.*, 1972] all show that in running speech the dimension of laryngeal timing is a powerful differentiator of homorganic consonants.

I cannot refrain from alluding to two serious misunderstandings of our concept of VOT. In a purported demonstration of the unimportance of VOT for English initial stops, Winitz *et al.* [1975] manipulate the onset of voice timing as a completely independent variable. Thus, 'VOT' values were altered in real-speech recordings in such a way as to yield improbable and even impossible temporal combinations and sequences of voice pulsing and aspiration. Using the resulting 'syllables' as stimuli in perception tests, they claimed to show that aspiration is the major cue to voicing distinctions, while VOT is a secondary cue. Clearly, these investigators have not grasped the central point that VOT is a laryngeal dimension with a complex set of intersecting, over-

[5] After all, even chinchillas have been trained to perceive VOT differences [Kuhl and Miller, 1975].

lapping or even discrete acoustic cues. To take, for example, an original English /du/ and move the consonant burst back so that there is a silent gap of 35 msec between it and the onset of voicing [WINITZ *et al.*, 1975; fig. 1] and say that this is the equivalent of a VOT value of +35 msec is simply untenable. An honest use of our concept and test thereof would reveal that such a value of VOT would include turbulent excitation of the upper formants and attenuation of the first formant. Now these authors have of course the perfect right to tease out any of the acoustic cues associated here with laryngeal timing, and perhaps others not yet mentioned, and test the perceptual efficacy of any one of them, as has been done, for example, for the completion of formant transitions before or after the onset of voicing by STEVENS and KLATT [1974][6] and the role of fundamental frequency by HAGGARD *et al.* [1970] and FUJIMURA [1971]. Although I readily concede that our terminology needs elaboration to cover the separate acoustic aspects of laryngeal timing [SCULLY, pers. commun.], it hardly behooves other investigators to cite us in denigrating VOT without reading closely to see that we mean much more than the mere timing of voice pulsing as a feature orthogonal to other consequences of laryngeal timing.

The other recent instance of misunderstanding I have in mind is a study of voicing and aspiration in Hindi stop consonants by BHATIA [1976]. The author somehow interprets the work on VOT by LISKER and me [1964] and on the related matter of the size of glottal opening by KIM [1970] to predict the neutralization of aspirated and unaspirated stops in final position. To the extent that certain statements by KIM may be vulnerable to BHATIA's [1976] criticism, I have no wish to enter into the argument; nevertheless, degrees of glottal opening seem clearly relevant to the final distinctions in Hindi. Beyond that, since we would argue that except for the special states of the glottis required for such features as murmur and creak, the degrees of glottal opening needed for voicing distinctions, including voiceless aspiration, go hand in hand with laryngeal timing, I must protest that here too an investigator has failed to grasp the point that VOT is a utterance-initial manifestation of the more general phenomenon of laryngeal timing. Indeed, one could go further and argue reasonably that word-final aspiration is an instance of VOT. Consider that in an English word like *potato* the unstressed first syllable is likely to have no voicing at all; that is to say, it is

[6] LISKER [1975] clarifies the matter in experiments in which he pits a literal interpretation of VOT against 'voiced transition duration'.

completely aspirated so that VOT proper does not take place until well after the beginning of the second syllable. The result is a voiceless vowel in the first syllable. This is a case, if you will, of a voicing lag so extreme as to deprive a whole syllable of voiced excitation. To produce aspiration in final position, it is necessary to release the stop, thus articulating an unstressed additional syllable (or perhaps 'pseudosyllable'). This additional unstressed 'syllable' includes a noise-excited vowel appropriate to the vocal tract configuration of the moment. BHATIA's [1976, p. 73] remarks on the predictive powers of phonetic theories are quite gratuitous!

It must not be supposed, one early critic notwithstanding [KIM, 1965], that we have ever claimed that even in utterance-initial position the dimension of laryngeal timing will explain every distinction of homorganic consonants that apparently involves laryngeal features of one sort or another [LISKER and ABRAMSON, 1964, 1971, 1972; ABRAMSON and LISKER, 1970b]. VOT may be said to distinguish the voiced aspirated (murmured) stops of such languages as Hindi and Marathi from voiceless stops but certainly not from the voiced unaspirated stops. Here VOT intersects with the kind of glottal opening that permits weak but audible phonation to occur with simultaneous turbulence [HIROSE *et al.*, 1972]. For the three stop categories of Korean, VOT gives mixed results [LISKER and ABRAMSON, 1964]. In word-initial position, two of the categories show a fair amount of overlap although the two of them are well separated from the third. These data taken with the rather complicated response patterns of perceptual experiments with VOT [ABRAMSON and LISKER, 1972] led us to conclude that the timing of glottal adjustments relative to supraglottal articulation does contribute to the Korean distinctions but that there must be another dimension that works with VOT in distinguishing the stop categories. The latter conclusion has been borne out by fiberoptic and electromyographic studies [KAGAYA, 1974; HIROSE *et al.*, 1974].

Shifts in three extralaryngeal features are commonly adduced in descriptions of the voicing distinction: the volume of the supraglottal tract, stop closure duration in medial position, and vowel duration before a final stop. For phonation to be sustained during an occlusion of the supraglottal vocal tract, it is necessary to prevent equalization of transglottal air pressure. ROTHENBERG [1968, p. 91] calculates that without any special adjustment this equalization would occur in 4 msec which would allow only one or two glottal oscillations. With passive expansion of the pharyngeal walls, voiced closures could be accommodated up to 20–30 msec [pp. 93–94]. Active expansion of the pharynx, according to ROTHENBERG's [pp. 94–99] calculations might give voiced closure durations of 80–90 msec. The even longer voiced closure durations often observed [LISKER and ABRAMSON, 1964] might be explained by incomplete velopharyngeal closure [ROTHENBERG, 1968, pp. 99–106][7].

Expansion of the pharynx during voiced occlusions has been observed by a number of investigators, at least for citation forms. Apparently, because of a conviction that English voiced stops are 'lax' and voiceless stops, 'tense', some of them, e.g. PERKELL [1969], assumed that the pharyngeal walls expanded passively to help maintain the transglottal airflow for voicing, while the walls were tensed to prevent voicing for the voiceless stops. Electromyographic examination of the relevant musculature [BELL-BERTI and HIROSE, 1975; BELL-BERTI, 1975] reveals that one cannot predict for a given subject whether active or passive control, or some combination of the two, will be exercised for variations in the volume of the supraglottal cavity for voicing distinctions in English. The feature of pharyngeal expansion is linked with laryngeal timing, yet it may be independent. This is not known. For that matter, we do not know how reliable the feature of pharyngeal expansion itself is in running speech.

For some time [LISKER, 1957], it has been known that spectrograms of English medial voiceless stops before unstressed syllables show longer closure durations than do voiced stops and that manipulation of this feature, providing that no voiced pulsing is present during the closure, furnishes a sufficient cue for the perception of the voicing distinction. Whether this feature is independent or somehow has a dependency relationship with laryngeal timing is not known at this time. Comparison of closure durations across all principal environments, using oral air pressure traces [LISKER, 1972], shows that this feature is really reliable only in medial post-stressed position and thus much less useful as an index to the voicing distinction than is laryngeal timing.

The final non-laryngeal feature to be considered here is the well-documented observation that in English and some other languages vowels preceding final voiced consonants are longer than those preceding final voiceless consonants. This durational difference is perceptually relevant [DENES, 1955; RAPHAEL, 1972]. One attempt [HALLE and

[7] See, e.g., data for Sindhi [NIHALANI, 1975].

Stevens, 1967] has been made to tie this feature directly to the laryngeal control needed to maintain voicing during consonant closure. Since, however, voicing distinctions in final position are likely to be characterized by differences in laryngeal timing, namely voice offset time, the question arises as to whether the concomitant difference in vowel duration is completely independent of laryngeal timing.

Languages of the world make extensive use of the timing of the valvular action of the larynx relative to supraglottal articulation to distinguish classes of consonants. Certain non-laryngeal features accompany laryngeal timing, but it remains to be determined whether any of them are controlled by the same mechanism. Laryngeal timing underlies a complex set of interrelated acoustic features any one of which may have perceptual efficacy. The total set varying rather predictably with changes in laryngeal timing has differentiating power in speech perception. The focusing of attention for many years on utterance-initial position, reflected in the widely used term VOT, seems to have led some investigators to fail to understand VOT and its acoustic complexity as a positional manifestation of the more general phenomenon of laryngeal timing.

References

Abramson, A. S. and Lisker, L.: Voice onset time in stop consonants: acoustic analysis and synthesis. Proc. 5th Int. Congr. Acoust., Paper A51 (Thone, Liège 1965).

Abramson, A. S. and Lisker, L.: Discriminability along the voicing continuum: cross-language tests. Proc. 6th Int. Congr. Phon. Sci., Prague 1967, pp. 569–573 (Academia, Prague 1970a).

Abramson, A. S. and Lisker, L.: Laryngeal behavior, the speech signal and phonological simplicity. Actes 10e Congr. Int. Ling., Bucarest 1967, IV, pp. 123–129 (L'Académie, Bucarest 1970b).

Abramson, A. S. and Lisker, L.: Voice timing in Korean stops. Proc. 7th Int. Congr. Phon. Sci., Montreal 1971, pp. 439–446 (Mouton, The Hague 1972).

Abramson, A. S. and Lisker, L.: Voice-timing perception in Spanish word-initial stops. J. Phon. *1:* 1–8 (1973).

Bell-Berti, F.: Control of pharyngeal cavity size for English voiced and voiceless stops. J. acoust. Soc. Am. *57:* 456–461 (1975).

Bell-Berti, F. and Hirose, H.: Palatal activity in voicing distinctions: a simultaneous fiberoptic and electromyographic study. J. Phon. *3:* 69–74 (1975).

Bhatia, T. K.: On the predictive role of the recent theories of aspiration. Phonetica *33:* 62–74 (1976).

Cooper, F. S.; Sawashima, M.; Abramson, A. S., and Lisker, L.: Looking at the larynx during running speech. Ann. Otol. Rhinol. Lar. *80:* 678–682 (1971).

Denes, P.: Effect of duration on the perception of voicing. J. acoust. Soc. Am. *27:* 761–764 (1955).

FUJIMURA, O.: Remarks on stop consonants. Synthesis experiments and acoustic cues; in HAMMERICH, JAKOBSON and ZWIRNER Form and substance. Phonetic and linguistic papers presented to Eli Fischer-Jorgensen, pp. 221–232 (Akademisk Forlag, Copenhagen 1971).

HAGGARD, M.; AMBLER, S., and CALLOW, M.: Pitch as a voicing cue. J. acoust. Soc. Am. *47:* 613–617 (1970).

HALLE, M. and STEVENS, K. N.: On the mechanism of glottal vibration for vowels and consonants. Q. Prog. Rep. Res. Lab. Electron. MIT *85:* 267–271 (1967).

HIROSE, H.; LISKER, L. and ABRAMSON, A. S.: Physiological aspects of certain laryngeal features in stop production. Status Rep. Speech Res. Haskins Labs *31/32:* 183–191 (1972).

HIROSE, H.; LEE, C. Y., and USHIJIMA, T.: Laryngeal control in Korean stop production. J. Phon. *2:* 145–152 (1974).

KAGAYA, R.: A fiberscopic and acoustic study of the Korean stops, affricates and fricatives. J. Phon. *2:* 161–180 (1974).

KIM, C. W.: On the autonomy of the tensity feature in stop classification (with special reference to Korean stops). Word *21:* 339–359 (1965).

KIM, C. W.: A theory of aspiration. Phonetica *21:* 107–116 (1970).

KUHL, P. K. and MILLER, J. D.: Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. Science *190:* 69–72 (1975).

LIBERMAN, A. M.; DELATTRE, P. C., and COOPER, F. S.: Some cues for the distinction between voiced and voiceless stops in initial position. Lang Speech *1:* 153–167 (1958).

LISKER, L.: Closure duration and the intervocalic voiced-voiceless distinction in English. Language *33:* 42–49 (1957).

LISKER, L.: Stop duration and voicing in English; in VALDMAN Papers in linguistics and phonetics to the memory of Pierre Delattre, pp. 339–343 (Mouton. The Hague 1972).

LISKER, L.: Is it VOT or a first-formant detector? J. acoust. Soc. Am. *57:* 1547–1551 (1975).

LISKER, L. and ABRAMSON, A. S.: A cross-language study of voicing in initial stops: acoustical measurements. Word *20:* 384–422 (1964).

LISKER, L. and ABRAMSON, A. S.: Stop categorization and voice onset time. Proc. 5th Int. Congr. Phon. Sci., Münster 1964, pp. 389–391 (Karger, Basel 1965).

LISKER, L. and ABRAMSON, A. S.: Some effects of context on voice onset time in English stops. Lang. Speech *10:* 1–28 (1967).

LISKER, L. and ABRAMSON, A. S.: The voicing dimension: some experiments in comparative phonetics. Proc. 6th Int. Congr. Phon. Sci., Prague 1967, pp. 563–567 (Academia, Prague 1970).

LISKER, L. and ABRAMSON, A. S.: Distinctive features and laryngeal control. Language *47:* 767–785 (1971).

LISKER, L. and ABRAMSON, A. S.: Glottal modes in consonant distinctions. Proc. 7th Int. Congr. Phon. Sci., Montreal 1971, pp. 366–370 (Mouton, The Hague 1972).

LISKER, L.; ABRAMSON, A. S.; COOPER, F. S., and SCHVEY, M. H.: Transillumination of the larynx in running speech. J. acoust. Soc. Am. *45:* 1544–1546 (1969).

LISKER, L.; SAWASHIMA, M.; ABRAMSON, A. S., and COOPER, F. S.: Cinegraphic observations of the larynx during voiced and voiceless stops. Status Rep. speech Res. Haskins Labs. *21/22:* 201–210 (1970).

NIHALANI, P.: Velopharyngeal opening in the formation of voiced stops in Sindhi. Phonetica *32:* 89–102 (1975).

PERKELL, J. S.: Physiology of speech production: results and implications of a quantitative cineradiographic study (MIT Press, Cambridge 1969).

RAPHAEL, L. J.: Preceding vowel duration as a cue to the perception of voicing of word-final consonants in American English. J. acoust. Soc. Am. *51:* 1296–1303 (1972).

ROTHENBERG, M.: The breath-stream dynamics of simple-released-plosive production. Biblthca phonet., No. 6 (Karger, Basel 1968).
SAWASHIMA, M.; ABRAMSON, A. S.; COOPER, F. S., and LISKER, L.: Observing laryngeal adjustments during running speech by use of a fiberoptics system. Phonetica *22:* 193–201 (1970).
STEVENS, K. N. and KLATT, D. H.: Role of formant transitions in the voiced-voiceless distinction for stops. J. acoust. Soc. Am. *55:* 653–659 (1974).
WINITZ, H.; LARIVIERE, C., and HERRIMAN, E.: Variations in VOT for English initial stops. J. Phon. *3:* 41–52 (1975).

Prof. ARTHUR S. ABRAMSON, Department of Linguistics, University of Connecticut, *Storrs, CT 06268* (USA)